Skoltech
Skolkovo Institute of Science and Technology

Thesis Changes Log

Name of Candidate: Konstantin Gubaev
PhD Program: Materials Science and Engineering
Title of Thesis: Machine-learning interatomic potentials for multicomponent alloys
Supervisor: Prof. Alexander Shapeev

Date of Thesis Defense: 02 October 2019

The thesis document includes the following changes in answer to the external review process.

Apart from the questions discussed below, some critical and valuable comments were considered in the thesis, along with typos and minor remarks. I am grateful to all the reviewers for their time and efforts spent on reading and understanding my thesis.

Reviewer: Dr. Alexander Levchenko

1) Please clarify how magnetism was treated in Co-Nb-V and Al-Ni-Ti alloys.
p.64, p.68
To consider magnetic properties of Co(Ni) present in this alloy, we initialized calculations with parallel magnetic moments assigned to Co(Ni) atoms, and zero magnetic moments assigned to Nb and V (Al and Ti) atoms, thus searching for a ferromagnetic ground state of a certain structure.

2) Please report distribution of errors for your application examples, including maximum errors, not only mean errors.
p.63,p.64,p.69
-he energy MAE, RMSE and max. error measured on the training sets:
Cu-Pd:
1.9 meV/atom, 2.3 meV/atom, and 10.1 meV/atom respectively,
Co-Nb-V:
6.2 meV/atom, 8.1 meV/atom, and 29 meV/atom respectively,
Al-Ni-Ti 1st round:
18 meV/atom, 27 meV/atom, and 91 meV/atom respectively,
Al-Ni-Ti 2nd round:
7 meV/atom, 9 meV/atom, and 24 meV/atom respectively.

3) Mention tight-binding and other quantum-mechanical force fields.
p.19-20
There are approaches which search the solution of (11) in a form of a

linear combination of atomic orbitals, e.g., Hartree-Fock, post-Hartree-Fock methods. Among these are tight-binding model which treats electrons as "belonging" to atoms and uses linear combination of atomic orbitals to determine the electron energy levels. The opposite (in a sense of binding toatoms) approximation - nearly free electron model (modification of a free electron model) treats electrons as a gas with only weak interaction with ions, which allows for correct prediction of many features of the electronic structure, especially in metals, when outer electrons are essentially delocalized. Hybrid functionals in DFT (Section 2.3) also employ predictions by Hartree-Fock theory, and they are capable of calculating many chemical systems with high accuracy, introducing corrections to the exchange-correlation functionals (Section 2.3). Apart from relatively "heavy" methods like DFT hybrid functionals or, etc., quantum Monte Carlo (which operates with exact manybody wavefunction and treats quantum effects directly) some QM models are semi-empirical and adjusted for relatively fast calculations, providing approximate yet often accurate predictions for large systems, which evaluation with more costly methods would be prohibitively long; e.g., linear scaling DFT, which incorporates some screening of interatomic interaction leading to O(N) cost scaling with the number of atoms, versus O(N^3) for conventional DFT.

4) There is no clear motivation for developing "yet another" approach (MTP) machine-learning potential. Please motivate clearly why your approach is better than the approaches currently developed by other groups.

p.24

Even with all variety of ML methods developed for atomistic modeling in recent years, this only reflects the immaturity of the present MLIPs field and continuous attempts to search for more generic, more accurate and more robust approaches. Of course, all the ML methods enlisted in this section suit for certain problems they were developed for, however they possess certain drawbacks, motivating the development of yet another ML approach, e.g.: the GAP [81] requires much more computational time to achieve the same accuracy compared to the single-component MTP (see [76]), NN-based models usually require a lot of training data (see [32]), cluster expansion is limited to lattice-based structures. From my personal point of view, the drawbacks of contemporary MLIPs are a consequence of one of the following reasons: the concept of a MLIP can be too \physically-motivated" (e.g, EAM potentials) or vice-versa, largely inspired by ML and thus not well appropriate for describing the physical (atomistic) systems. Physically inspired models can suffer from lack of flexibility in their functional forms, thus failing to capture complex (features of interatomic interactions not reflected in \physics" of their design. On the other hand, ML-based models can neglect or not fully account for the origin of the data they are fit to, which results in excessive attempts required to capture basic physics phenomena (e.g., non-differentiable ML models like random forest predicting energy but not forces, which are essentially the derivatives of the energy). From this perspective the MTPs incorporate a balance between physically inspired descriptors and flexible polynomial functional form (see Section 3.4) making approximation of PES accurate with relatively small number of parameters [32] and allowing for making more accurate MTPs by employing more complex functional form. In addition, the active learning approach (Section 3.6) developed for MTPs solves the sampling problem (Section 3.2), which arises in any simulation involving ML models and can be more difficult to overcome, then the problem of accurate MLIP fitting.

5) Please motivate better why one needs to study the considered alloy systems.

p.63,p.64,p.68

- We chose the Cu-Pd system because the structure of both pure Cu and Pd is fcc, while the stable equimolar CuPd structure is a bcc derivative structure. This system is a good test of whether or not our MTP-based model is able to simultaneously handle multiple lattice types.
- We next test our algorithm on constructing a convex hull for the ternary Co-Nb-V system in the region where the concentration of Co is 50% or more. The choice is motivated by the several Co-Nb and Co-V binaries present in this region of the phase diagram, which our approach should predict.
- Finally, we applied our algorithm to the Al-Ni-Ti system. This system is well-studied and has many known ternary structures, some of which have over 20 atoms in the unit cell, therefore we considered this system a good test for our approach.

6) Eq. 11 (Born-Oppenheimer approximation) and others: discuss spin

p.19

Though no electronic (or nuclear) spins were considered in the derivation of (11), they can be included in (11) directly as its derivation takes into account only spatial degrees of freedom.

p.21 (spins in density functional theory)

To account for spin degrees of freedom, similarly to (2.3) spin-up and spin-down densities can be introduced, calculated from corresponding spin-up and spin-down wave functions.

7) "However, in practice it is unsolvable in principle" - clarify that for some choices of F(theta) it is solvable (namely linear; maybe some others?)

p.29

Exact solution of the optimization problem can be found only in case of polylinear or polyquadratic L(theta). i.e., the minimizing functional has linear/quadratic dependence on the internal parameters.

8) "Their typical values are: Ce = 1; Cf = 0.01 A^2; Cs = 0.001 A^6 ..." - explain where these typical values come from

p.39

The numbers from (21) come from practice and reflect the condition of optimal relative importance of the energy, forces, and stresses factors in the fitting procedure. I.e., too small or too large value of any coefficient from (21) will result in neglecting/overprevailing of a certain factor among the others.

9) "Though such a shallow ground state is typically not significant beyond academic interest" - clarify your criterion for shallow/non-shallow ground state

p.63

The value of 0.5 meV is comparable to thermal energy of one atom at 6K temperature, thus this ground state can easily be escaped due to thermal fluctuations in any realistic scenario.

Reviewer: Dr. Ferenc Tasnadi

10) I suggest to elaborate/extend the last paragraph of section 2.4 as a state-of-the-art and put the MLIP approach into a context of other methods in crystal structure prediction with applications and limitations.

p. 24
See question 4 where the last paragraph of the section 2.4 is provided.

11) On page 33 you write that validation can be done with predicting “high-level” quantities. Fitting the model with quantum mechanical total energies and the predict derivatives of energy, such as elastic constants, phonons can be inaccurate. Is there any argument/motivation to underline your statement?

Comment: the purpose of “validation” is to somehow check the performance of the potential apart from the fitting quality (which is more related to the errors on the training set). To that end, fitting energies and predicting their derivatives ofc. can be inaccurate, but performing training and validation with the data of different origin is a factor can be seen as positive.

p.36

A common way of checking the performance of the potential outside the training set is using a validation set (see Section 3.1). However, it is not always possible to construct a validation set. It requires configurations relevant to the simulation of interest with provided ab initio energies/forces/etc., which is not always possible. The reasons can be: the potential can be used for large supercell calculations, for which ab initio calculations are not possible, or the configurations from the simulation are not known in advance, because the trajectory of the MD is known only during actual simulation. Therefore, sometimes validation of the potential is done versus more “high-level" quantities like vacancy migration energy barrier, elastic moduli or heat capacity: the potential can be used to calculate some of these quantities with further comparison to their known values. This allows for estimation of the potential quality: if the potential reproduces some quantities well, it can be expected that it will properly predict other properties of interest. Obviously this is not a systematic criterion as opposite to the active learning approach, see Section 3.6.

12) Is it possible to extend the ordinary 3D space and define a descriptor for magnetic materials? (Introduce magnetic implementation of MTPs)

p.46

Despite the magnetic moments of the atoms are neglected in the current implementation, they can be included in the current implementation by adding a dependence of the radial functions f from (26) on the magnetic moments of atoms in each pair participating in the tensor moment construction. The below formula provides generalized radial functions expression for a case of collinear magnetic moments:

$$f_\mu(\rho, z_i, z_j, m_i, m_j) = \sum_k c^{(k)}_{\mu,z_i,z_j} Q^{(k)}(\rho) \sum_{k_1,k_2} c^{(k,k_1,k_2)}_{\mu,z_i,z_j} Q^{(k_1)}(m_i) Q^{(k_2)}(m_j), \qquad (27)$$

where (28)

$$Q^{(k)}(\rho) := T_k(\rho)(R_{\text{cut}} - \rho)^2.$$

Compare to radial functions of non-magnetic MTPs:

$$f_\mu(\rho, z_i, z_j) = \sum_k c^{(k)}_{\mu,z_i,z_j} Q^{(k)}(\rho), \qquad \text{where} \qquad (26)$$

$$Q^{(k)}(\rho) := T_k(\rho)(R_{\text{cut}} - \rho)^2.$$

In this implementation (27) the magnetic moments of atoms are treated as independent variables on which the potential energy depends, which adds two dimensions (as radial functions are constructed for pairs of atoms) to a space of independent variables on which f depends, and two additional sums in the expansion of this function through polynomials $Q^{(k)}$. In a general case of 3-d magnetic moments, it will be 6 additional dimensions and 6 sums over corresponding indexes.

13) Could you please explain the meaning of non-locality, because both Figure 22 and equation (32) in the thesis suggest rather the incompleteness of the models MTM1 and MTM2 instead of non-locality. If both MTM1 and MTM2 were complete then they would provide the same result, or?

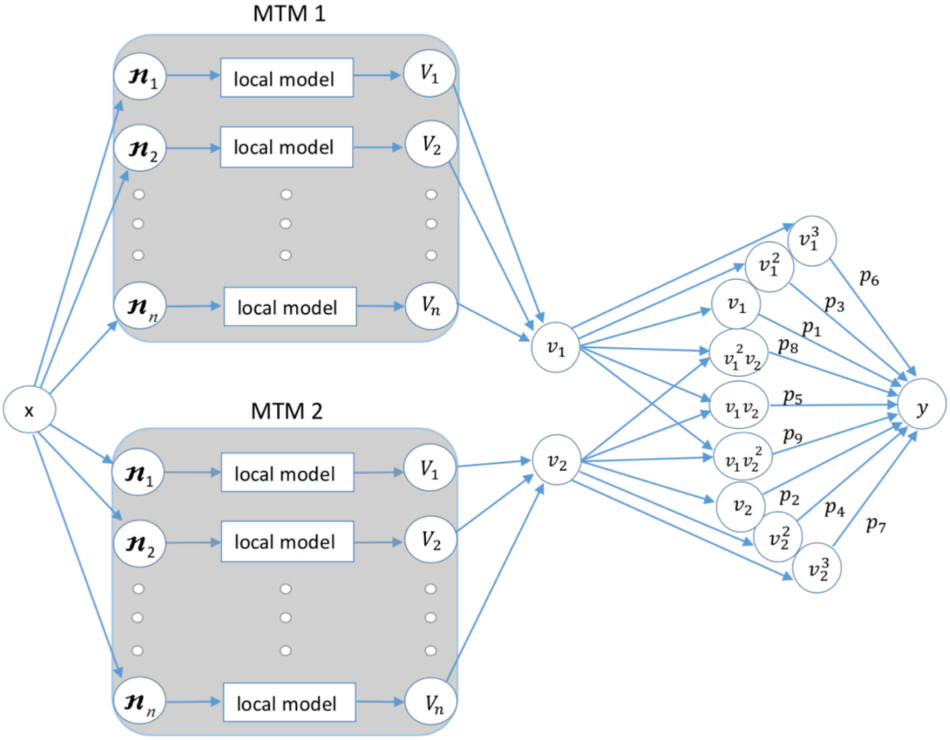

Figure 22: The nonlocal model, nlMTM, constructed from the two local models. $p_1, \ldots, p_9$ are the additional fitting parameters.

p.84-85

Taking into account relatively poor performance of conventional MTMs for HOMO and LUMO molecular quantities (see 3) we decided to include nonlocal effects by introducing two different local models v_1 and v_2 (each with its own set of parameters) and let:

$$v_1 = \sum_i V_1(\mathfrak{n}_i), \qquad \text{and} \\ v_2 = \sum_i V_2(\mathfrak{n}_i). \tag{31}$$

We then define the nonlocal model *nlMTM* in the following form

$$F_{\text{nl}}(x) = p_1 v_1 + p_2 v_2 + p_3 v_1^2 + p_4 v_2^2 + p_5 v_1 v_2 \\ + p_6 v_1^3 + p_7 v_2^3 + p_8 v_1^2 v_2 + p_9 v_1 v_2^2. \tag{32}$$

Molecular orbitals get occupied by electrons not independently of each other, hence we assume a nonlinear dependence of the answer on the local features (32).
E.g., on Figure 23 the neighborhoods of atoms 1 and 4 do not intersect, therefore neither pair interaction 1-4 nor triple interactions 1-2-4 or 1-3-4, etc. cannot be approximated by the local MTM model.

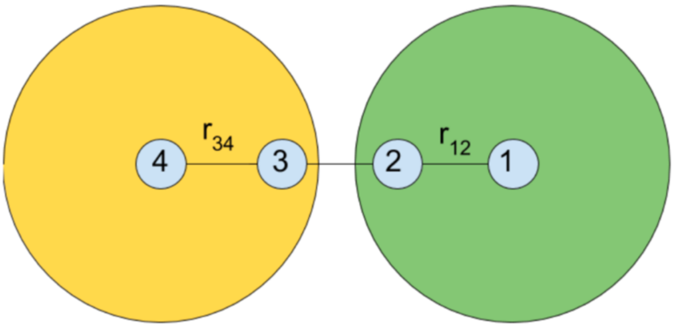

Figure 23: Neighborhoods of atoms 1 and 4 do not intersect. Therefore mutual interaction of atoms 1 and 4 cannot be taken into account by a local model with current cutoff radius.

As opposite, following the formulas (32) and (31), the approximated nonlocal quantity will include, e.g., the following terms:

$$F_{\text{nl}}(x) = p_5 v_1(r_{12}, r_{34}, \ldots) v_2(r_{12}, r_{34}, \ldots) + \ldots \\ = p_5(V_1(r_{12}) + \ldots) * (V_2(r_{34}) + \ldots) + \ldots \quad = \underbrace{p_5 V_1(r_{12}) * V_2(r_{34})}_{f(r_{12}, r_{34})} + \ldots \tag{33}$$

The underbraced summand in (33) depends on both r12 and r34. Similarly, after expanding all v_1's and v_2's in (32) the expression for $F_{\text{nl}}(x)$ contains polynomials of rij mutually multiplied independently of the distance between the corresponding atoms and thus forms a basis for approximating nonlocal properties similarly to local model, see Section 3.4.

14) Can one use [the Figure with a single-component case from the Ref.[60]] with force error to validate the MaxVol approach to multicomponent systems, why?

p.51

Obviously, to some extent the extrapolation grade correlates with the error which MTP is expected to have for a certain configuration. Though, the authors emphasize that the essence of this criterion is measure of proximity of configuration to the training domain of a certain MTP. Nevertheless, it would be illustrative to provide such dependence for a case of MTP designed for a TiN binary system. The MTP used for this test has 50 parameters and configurations are taken from NPT simulations of B1 TiN at 300K.

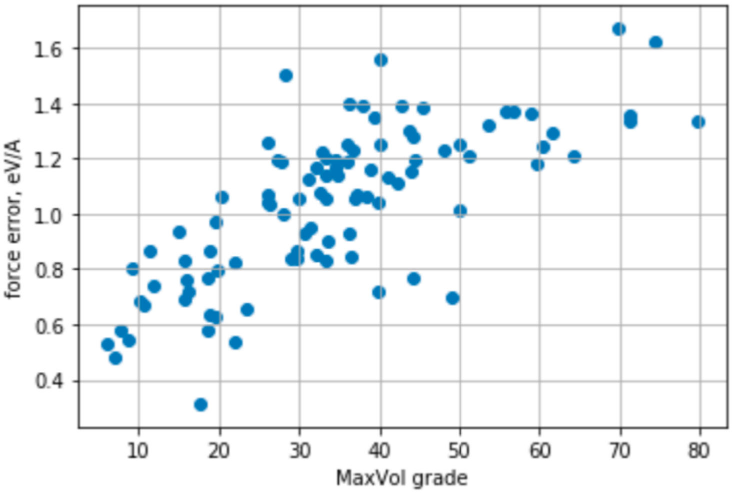

Figure 8: The illustration of a connection between extrapolation grade and the absolute force error of MTP.

15) Can you explain how to choose DFT calculation parameters for the best “accurate prediction/computational demand” ratio? Do you have any strategy? Can you consider to add some description to the thesis?

Comment: My strategy is rather to have robust and accurate enough calculation even if it is not very computatutationally fast, and I use recommendations from VASP website and from more experienced users. I essentially skip providing comments on choosing the settings, as this is a separate big discussion, falling beyond the scope of the thesis.

p.63,p.64,p.69

- PAW_PBE GGA potentials
- Automatic K-mesh with KSPACING=0.15
- Energy cutoff: Cu-Pd 500meV, Co-Nb-V and Al-Ni-Ti 400meV

16) What type of relaxation approach has been used?

p.59

To perform relaxation we treat energy of configuration as a function of atomic positions and lattice vectors. As we know the derivatives of energy function w.r.t. this variables (calculated from forces and stresses), we use BFGS optimization algorithm (as in Section 3.1) for minimizing energy, which simultaneously provides zero forces and stresses.

Reviewer: Dr. Biswanath Dutta

17) How important in the choice of the initial set of configurations for the quality of MTP?

p.54

It is assumed that the starting training set is empty and is only composed algorithmically as is described below. This allows for excluding the human factor on this stage and therefore making the procedure more generic and more automatized. One can, however, still start from some initial set of configurations to end the procedure faster with ssome final training set as the result (which is still more or less the same regardless of the initial training set). To benefit from the active learning approach one should not provide too much configurations from the beginning, to leave the space for new incoming configurations. While the size of the final training set would be about 2m, where m is the amount of parameters in the MTP, we recommend to take not more than roughly m=2 configurations for initialization.

18) Will it be possible to see a phase transition if the MTP is trained with data from only one of the phases?

p.59

The active learning approach proposed in this section allows for "exploring" the unknown parts of the configurational space during a simulation: e.g., if an atomistic system during an MD is in the state preceding a phase transition, the system will evolve towards the new phase with increasing number of extrapolative configurations occurring. Whether the phase transition will be reproduced correctly depends on many factors, among which the accuracy of the MTP fitting is crucial. However in principle, atomistic simulations with active learning are designed to be suitable for such problems without the *a priori* knowledge about the studying system: the only source of information which "guides" the simulation is the QM values for energies, forces and stresses.

19) In the present thesis the machine-learning approach previously developed for single-component systems has been extened to multiple component alloys. Can you describe what were the key challenges you faced and how did you handle those challenges?

p.46

While training of a single-component MTP means solving the overdetermined system of linear algebraic equations (a single-component MTP is linear w.r.t its parameters [76]), which is purely procedural and straightforward task, training a multicomponent MTP requires solving the non-linear optimization problem in high-dimensional space. For being able to use the gradient-based optimization methods like gradient descent one needs to compute the first (at least) derivatives of the loss function (19) w.r. its parameters, which takes much more computational effort then calculating just a loss function value. Therefore I have implemented a back-propagation algorithm for derivatives calculation which calculates all the loss function derivatives w.r.t MTP coefficients (which are hundreds) only 4.8 times slower than the actual loss function calculation. Another important question arising during non-linear optimization is choosing the optimization method.

20) Can MTPs handle short-range ordering?

p.56

MTPs cannot predict the long-range ordering in materials as they are essentially local (see Section 2.1). Due to the same reason MTPs by their nature learn the short-range order by intrinsically assigning different energy contributions to different atomistic environments. This way, more energetically favourable short-range orderings will be distinguished among all possible orderings.

Reviewer: Dr. Nikolay Gippius

21) The chemistry application appear to predict properties of equilibrium molecules. Are MTPs able to predict chemical reactions?
p.86
As MTPs allow calculating forces due its differentiable analytical form (Section 3.3) they are in principle capable of describing chemical reactions. However, this have not been explicitly tested for any particular reaction.

22) MTP compute the interaction energy based on the atomic coordinates. What happens if there are atomic charges and spins present in the system?

Charges are not treated explicitly, comments added on p. 46:

The only means by which the MTPs can describe charged systems are screening of electrostatic interaction, which makes the locality assumption (2) applicable. In the case of essentially non-local electrostatic interaction the acharges should be included explicitly within hypothetical non-local implementation.

Spins: see question 12, p.46

23) In what sense do MTP extrapolate?
p.49
MTP "extrapolates" when configuration does not belong to the interior of the configurational space, covered by the training set. This essentially means proximity of the configuration to the training set of MTP. Definition and detailed description of the extrapolation grade concept is given in section 3.6.1. The approach we propose is based on the so-called D-optimality criterion, which provides a numerical criterion for “novelty" of a certain configuration for interatomic potential (namely, MTP).